Computers & composition in the US Government—past, present & future

E. R. Lannon

The US Government is a major publisher as well as the largest single user of electronic data processing equipment in the world. In 1964 printing costs, exclusive of authorship, were $254,000,000. Approximately 20 per cent of the publications produced for the sum stated were composed in part or in whole on one or more of the 2,600 electronic data processing systems installed by the Government. These computers use 71,000 employees and involve the expenditure of over $2,000 million annually. Thus it is that computers and composition are of considerable interest to us.

Beginnings of use of computers as composing devices

The first of our many business-type digital computers was the first Univac I marketed. This computer was installed at the Bureau of the Census in March of 1951. Its output printer was a modified typewriter which operated at the speed of 12 characters per second from magnetic tape input. Soon after delivery it was pressed into service to prepare camera copy on statistical studies. It was not until 1955 that the first high-speed line printer was added to the Univac I (by this time the Burroughs 205 and IBM 650 with modified tabulators as output printers were being installed in other units of the Government). The 600 lines per minute printer of the Univac I is quite similar to Remington Rand output printers of today with the exception of speed. The outstanding advantages of using high-speed printer copy for reproduction were the savings in time and manpower. In earlier censuses, tables had to be posted by hand, reviewed, typed (or sent to the printer for composition), proofread, and machine checked. For the 1960 censuses, the computer output was programmed so that it fell into columns on preprinted forms. The original copy was sent to a copy-preparation unit and a carbon to the appropriate subject specialist. The copy-preparation operations consisted of spraying to protect the copy from smearing, inserting preprinted page headings, page numbers, and table titles, and ruling column and heading lines. If no corrections were needed after subject review, the tables were usually ready at this point for reproduction by photo-litho. The electronic equipment was considered so accurate as to eliminate the need for the proofreading and machine checking required for typewritten or typeset copy.

The high-speed printer copy for diary review and other internal use was produced at the rate of 600 lines a minute, but publication copy was produced at 400 lines a minute to ensure the clarity of the imprint and satisfactory alignment.

Although the advantages of using the high-speed printer output of Univac I for camera copy for offset reproduction far outweighed the disadvantages, this technique also raised some problems:

Loss of space: The high-speed printer type printed 10 characters or digits to the inch as compared with 12 to the inch for material prepared on a typewriter with elite type. In addition, the overall width of the printer output was 120 units in 130 spaces, while the standard size page for census reports ($9\frac{1}{8}'' \times 11\frac{3}{8}''$) could accommodate 150 spaces of high-speed printer copy (or 180 spaces of typewritten copy). To avoid a severe loss of space and to increase the amount of data published on each page, the full 150 spaces on each page were used by having two 'passes' through the printer employing a technique originally developed for the 1958 economic censuses. While this required extreme care in the alignment of the two portions of each line, the cost in printer time and manpower was justified by the more efficient use of the space on each page.

Another factor that tended to cost space was that, unlike a typewriter which could provide half spaces between lines, the high-speed printer furnished only full spaces between lines.

All-in-all, the use of the high-speed printer output as camera copy accounted for a loss of approximately 20 per cent of space per page.

Presentation problems: The high-speed printer output of Univac I was limited to capital letters. The printer has no lower case, no bold face, and no italics.

In addition, the high-speed printer had no superior numbers to use to identify footnotes and only one symbol—the asterisk—for standard footnotes. To counteract this limitation, footnote symbols were inserted manually on a typewriter with type matching that of the high-speed printer. Another peculiarity of the printer output was the large black imprints made by periods. In the case of leaders used in the stubs of statistical tables, the heavy imprint and uneven alignment of the periods did not meet presentation standards. The problem was solved by dropping every other period.

Unsatisfactory positioning of commas by the high-speed printer, which printed them midway between lines rather than on the line, led to the decision to eliminate the comma in numbers and to substitute a space in its place.

Indeed, even though monospaced there are still many types of publications composed on the computer whose bulk we cannot

reduce by the newer techniques which will be discussed later. In these instances faster speeds of composition will probably cause the transition to high-speed photo typesetting.

Today as stated earlier upwards of 20 per cent of all publications of the US Government are composed on the line printer. In many instances (e.g. long catalogue listings subject to fairly infrequent change) the computer's output is used as manuscript for re-keying.

Notwithstanding the speed of the computer's line printer as a composing device we find serious fault with it. Chief faults are:

Some sacrifices in quality and readability
The typical computer-printer or tabulating machine produces monospaced upper case characters, that is, evenly spaced capital letters. The result is a monotonous looking product with no typographic variation between captions and body. There is little reader appeal and the eye must work overtime to find what it wants from the printed page.

Increased bulk
Since each character produced by a computer-printer takes the same amount of lateral space the letter 'I', for example, is given the same space as the letter 'M'. Thus in monospaced type we can put one-third fewer characters on a page than we can with proportionally spaced type such as is produced on the Linotype and similar composing machines. Bulk further increases because the computer, of necessity, uses a constant vertical dimension for each character. In conventional typography, on the other hand, we can vary the vertical dimension by using different sizes of type for specific purposes. We estimate that the average book produced from computer print-out contains 40 per cent more pages than the average book produced by typesetting equipment.

Increased printing and binding costs
As a corollary of having fewer characters on a page and standard vertical dimensions, we have found that computer-printer output results in larger expenditure on platemaking, press time, paper, and binding. Added costs also mean increases in total production time. Even with the speed of the computer-printer, which we measure in hundreds of lines per minute, it is sometimes questionable whether any time is saved if the entire production cycle is considered.

Loss of some computer potential

Operation of a computer resembles an industrial assembly line. Raw material is put in, processed in the central processor, and then the finished product emerges as the output. In any comparison we must consider the relative speeds at which the three elements within the computer assembly line operate.

The conventional input devices, such as card readers and punched paper tape readers, operate at speeds which we can measure with the time unit of seconds.

The central processors are so fast that seconds are an unwieldy measure of speed. Increasingly, the nano-second (billionth of a second) is becoming the unit of measure for the central processor.

The line printer as contrasted to the central processor is slow and operation is again measured in 'seconds'.

These imbalances in the speeds of the input, processing, and output elements mean that the full potential of the central processor is not being utilized. As a result, off-line processing on small computers for input and output functions has become quite common since 1960. The use of small computers to perform the slow functions permits the large computer to work with faster input and output media such as magnetic tape, the reading and writing speed of which we measure in thousands of characters per second. Tape speeds are commercially available today at rates up to 360,000 packed numeric characters per second. For alphabetic work, tape speeds are available in the range of 180,000 alphameric characters per second.

Third generation computers, now being delivered, depart somewhat from the concept of off-line input and output processing by the use of multi-processing techniques. In multi-processing, several programs are handled simultaneously by a single computer. Even here with multiple input-output devices directly connected with the processing unit there are still speed imbalances.

Industry is hard at work on both the input and output sides of the computer to obtain better balances in speed and more efficient use of the computer. Optical scanning, either off-line or on-line, has much promise for reducing the costs of input preparation and input time. Some work has been done on high-speed outputs; however, the limitations previously expressed with respect to the line printers still exist.

MEDLARS (Medical literature analysis and retrieval system)
The genesis of Medlars was the Index Mechanization Project Medicine, US Public Health Services, of my Department. Many embarked upon in June of 1958 by the National Library of of the objectives of this project were directly associated with printing a major publication then called the *Current List of Medical Literature.* From this project came an improved input system and a high-speed composing system based on sequential punched cards. The Listomatic process was adopted and the old *Current List of Medical Literature* was renamed *Index Medicus* (this latter perhaps to prove that Latin still has a place in the world of medicine). As a publication system the new system was far superior in all respects to its predecessor. The Listomatic process suffered seriously from a total systems point of view in that the sequential card which carried the citation to be printed lacked the capability to tell much, in the space available, about the citation itself. For information storage and retrieval purposes the punched card, particularly one with a limited number of columns, could not do much. Before, however, discounting the punched card for smaller systems, current work on code compression should be reviewed. The Xerox Corporation, for example, is doing considerable work on card retrieval systems which carry reduced size images produced by Xerography on punched cards and other types of processable documents with code compression techniques being employed that permit a great

deal of information to be displayed in a very few columns. Thus it was that Medlars began.

In June of 1961 a feasibility study was initiated and a systems design produced by the General Electric Company under contract to the National Library of Medicine. The preliminary system design was converted into a detailed systems design and finally into an operational system which had the following objectives:
(1) Improve the quality and enlarge *Index Medicus,* and at the same time reduce the time required to prepare the monthly edition for printing from 22 to 5 working days.
(2) Make possible the production of other compilations similar to *Index Medicus* in form and content.
(3) Make possible for *Index Medicus* and other compilations, the inclusion of references derived from other sources, as well as from journal articles.
(4) Make possible the prompt (a maximum of two days) and efficient servicing of requests for special bibliographies, on both a demand and recurring basis, regularly searching up to five years of stored computer files.
(5) Increase the average depth of indexing per article by a factor of five, ie, 10 headings versus 2.
(6) Nearly double the number of articles that can be handled annually—from 140,000 at the outset to 250,000 in 1969.
(7) Reduce the need for duplicative total literature screening operations.
(8) Keep statistics and perform analyses of its own operations, to provide the information needed to monitor and improve system effectiveness.
(9) Permit future expansion to incorporate new and as yet not completely defined—and hence secondary—objectives.
Not only were these design objectives met with the publication of the July 1964 issue of *Index Medicus* but also in the process the world's fastest photo typesetting system, known to us as GRACE (Graphic Arts Composing Equipment), was developed.

As early as October of 1961 we found that design objectives with respect to output speed could not be met with conventional composing machines. Punched paper tape is one of the slowest types of output that can be produced on a general purpose computer—maximum speed 300 characters per second, and the fastest composing machines available to us at the time were in the range of 12 characters per second. Considering that we had to produce roughly 400 characters per second it would have been necessary to have upwards of 30 conventional composing machines available as well as two paper tape punches on-line to the computer operating simultaneously. Space and cost factors forbade this. Surprisingly, we also found that if we considered the total time cycle to the production of film for imposition even the legendary speeds of the line printers were outside the systems design. We had a choice to make: relax our speed requirements somewhat and sacrifice the features of typography desired or embark on a new technology. With considerable gall and many prayers we chose the latter course and a new era began in the integration of printing with electronic data processing.

Grace was accepted by us in May of 1964 and produced the July 1964 issue of *Index Medicus.* Grace is also used now to produce additional publications which are similar in format to *Index Medicus.* Gradually we will be producing more and more specialized indices with decreased emphasis on an omnibus index such as is *Index Medicus.* Grace produces a full page of *Index Medicus* on to film positives at a speed of a page per minute. The achieved throughput rate is therefore in the range of 300 characters per second. Grace has a repertoire of 226 characters. Type faces employed by us are 6-point Crown Regular and bold, 10-point Century bold and 14-point Techno-bold wide.

We do not hyphenate any of the material composed on Grace. We found that the effort in our case was not worth the gain. We also found in almost two years of publication and with a world-wide audience not one complaint about the lack of hyphens!

The Grace device was manufactured for us by Photon and has logic components specific to the language of the Honeywell 800 computer.

Paper tape conversion systems

In March 1963 the US Government Printing Office entered the field of typesetting by computer with the installation of a Linofilm photo typesetter, an Elektron linecasting machine, and a conversion device to translate seven level $\frac{1}{2}''$ magnetic tape to 15-channel Linofilm paper tape. At present 5 Linofilm photo units and 5 Elektrons are in use.

The computers feeding data to this system do so in a form more compatible to their speed potential than paper tape. We can output in excess of 50,000 characters per second on magnetic tape compared to only 100 characters per second on paper tape. 300 character per second punches are available but not generally used in general purpose data processing.

Through various computer programs the US Government Printing Office now produces a variety of publications from computer processed materials. *The Technical Abstract Bulletin* and its derivative publication the *US Government Research and Development Reports* are examples.

The bulk of the work being processed in this manner was already on the computer. For one-time jobs not already on the computer and for reasons relating solely to the processing of the data, we are somewhat unconvinced that gains in productivity to be achieved by automating 10 or 12 character per second devices make computer processing economic. Perhaps limitedly so in the United States with our relatively high costs for keyboard operation. Not so, however, in my view, where labour productivity is high, labour costs low and capital costs high. Where speed is the operative factor, the economics of the situation may differ. Present thinking is that the equipment now being used for this range of work will be displaced when the Electronic Composing System to be installed towards the end of 1966 becomes operative.

Electronic composing system

Shortly after my Department was authorized to proceed with the development of the Grace device in March 1962 by the Joint Committee on Printing of the Congress of the United States, the Public Printer of the United States was asked to institute a research and development programme to overcome the problems of printing from copy generated by automatic data processing equipment.

A committee known as the Electronic Printing Committee was formed of representatives of the executive branch of the US Government and representatives of the Government Printing Office, which for reasons now mainly historic is part of the legislative branch.

The committee, of which I am a member, worked long and hard to produce the specifications for what we believe is the most advanced typesetting system in the world. The specifications were released to the industry in 1963 and an award was made to the Mergenthaler Linotype Company in March 1964 for the installation of two high-speed units in 1966. Present schedules call for delivery of the system early in 1967.

The system has as its major objective the handling of those publications which are now being composed on the line printers of our 2,600 electronic data processing systems. We estimate that dependent upon the size of press run we can reduce printing and binding costs up to 40 per cent. We estimate also that on typical publications now composed on automatic data processing machines we can reduce production times also by up to 40 per cent. Even more substantial savings are possible in some cases. While we are achieving these rather dramatic savings in cost and time, we will produce more legible publications with decreased bulk and decreased drain on the computers being employed.

The electronic composing system consists of a general purpose computer which has been programmed with the master composition program, the electronic photo typesetting machines, and the necessary peripheral equipment such as film processors, tape converters, and input keyboards.

The electronic photo typesetting machine is capable of generating, displaying and positioning any symbol, in random order, and as specified by a control and information magnetic tape which has been generated for it by an electronic computer. The output may be photographic film, paper, or eventually short run plates, the quality of which in all respects will be at least equal to the quality standards of the graphic arts industry.

The maximum image size is 48 × 63 picas (8″ × 10½″). A continuous listing may be produced up to 10½″ wide. Film or paper may be loaded in 4″, 6″, 9″, 11″ or 12″ widths.

Speed

The speed range varies with character size. At 6-point (classified advertising size) speed will be 1,000 characters per second, at 8-point (newspaper text) 800 characters per second, at 10-point (normal book size) 620 characters per second. The machine has the capability of generating a variety of point sizes from each of 4 character grids. Thus each character can be displayed in 5, 6, 7, 8, 10, 12, 14, and 18-point size. These sizes cover all of the normal sizes used in Government printing.

As an example of the composing speed, a 4,000 character page set in 8-point type will be produced in less than 6 seconds or at the rate of approximately 600 pages per hour, including the half second necessary to advance the film at the completion of the page.

Character grids

A standard character grid (or job set) which is comparable to conventional job sets now used in the industry, will have a complete font of type including alphabetic, numeric, punctuation, and related graphic arts characters, in roman, italic, small caps, and bold. Grids will also be available to produce superior and inferior characters, as well as fractions and standard mathematical symbols.

A 256-character grid or set of type faces will always be available at electronic speed. Three additional 256-character grids are located in a turret and can be selected and placed automatically in half a second, giving a selection of 1,020 characters. One character on each grid is used for internal checking of the machine and is not available as a data character.

Principle of operation

The machine is housed in four separate cable-connected cabinets. The first contains the magnetic tape reader, a conventional magnetic tape drive as used by a computer. The second houses the tape reader control and control electronics. The third section contains the character generator. Finally, the page display cathode ray tube and film carrier complete the unit.

The character grid, previously mentioned, is optically projected to the face of a character generator tube. The photo cathode face of this tube changes the light image of the characters into electrons. Vertical and horizontal selection grids receive a signal from the control electronics and a single character emerges as a video signal.

Servo controls also receive information from the character generator to aid in the precise positioning of the character on the display cathode ray tube.

Let us follow a character through the photo typesetter from the magnetic tape to the final film output. The light source is constantly projecting the characters on the grid to the photo cathode. This is constant. There is no off-on condition. The operator pushes the start button and 2,048 characters (data and control) are read into the control buffer from the magnetic tape reader. The control electronics read the machine function codes such as point size or grid number, and the first character is ready to select.

The position of the first data character in relation to its place on the film page has been coded on the magnetic tape by a computer. The character code is decoded, and a signal is sent to

the character generator to open the gate for that character and let it emerge from the character generator as a video representation.

The video signal for the character is sent to the display cathode ray tube and positioning information is sent to the servo (control) cathode ray tube. The servo cathode ray tube and electronics position the character precisely and then send control information to the display cathode ray tube to display the character on the tube and through lenses to the film transport and photographic film.

As soon as a character is scanned the following character is released through the same process until a page is composed and a film advance signal is received to start a new page—all at the rate of approximately 1,000 characters per second.

With the proper signal in the magnetic tape, vertical and horizontal rules can be drawn in different thicknesses. A turn-page feature permits composition of broad pages (e.g., $10\frac{1}{2}''$ x $8''$).

The final output on film in page form will have an image resolution at least equal in quality to a line casting machine reproduction proof.

Master composition program
The electronic photo typesetting machine will not perform any operations that have not been written on to a magnetic tape by a general purpose computer controlled by the master composition program.

This computer program is a complex series of routines which provide the computer with instructions to perform every operation that a compositor must perform from setting galleys of type to page makeup. It will not be necessary for a Department programmer to have more than a general knowledge of this program which will include the instructions for preparing proper input.

In setting up varying types of composition, the master composition program must be given specifications for each individual job. These job instructions which are in all respects comparable to what must now be given to a compositor include the type face, type size, leading, line length, page width, page depth, running head location, and folio location.

These instructions are called page formats. A page format will be required for each variation in job instructions. As a format is developed for each new application, that format will be identified by code number and placed on a parameter tape file from which it may be called for future requirements of a similar nature. With instructions issued as a part of a programming manual, the programmer must incorporate in his input data magnetic tape the instruction codes to access the proper set of parameters into the central processing unit from the parameter magnetic tape. The master composition program then will have the necessary information to prepare a data and control magnetic tape for the photo composer.

Preparation of magnetic tape files
If the information which is stored on magnetic tape were prepared without knowledge that it would be used as input to the composition program, it is unlikely that it would contain all of the information needed by the composition program. An edit/insertion computer program must be written for that particular job. It is the function of the edit/insert program to identify and translate the file identifiers in that tape file to a set of codes which are recognized by the master composition program.

When information is processed specifically for the system, it will be possible in some cases to incorporate the necessary codes in the input data and produce output directly for the master composition program. The programmer familiar with the construction of the data file and the coded character data in that file is the person most qualified to write this edit/insertion program. The master composition program is being written for an IBM 360 computer. This program will later be adapted to meet the requirements of other computer systems as the need becomes apparent. This places no limitation on input to the system from any computer which is generating or maintaining data. Incompatible magnetic tape can be converted to the required half-inch magnetic tape.

The Lexical graphical composer system of the US Air Force

Stemming from work performed by the Mergenthaler Linotype Company on a project dealing with electronic translation of foreign language documents a graphics processor was developed which is now being made to work with the Electronic Composing System described above. Basically, this system permits the simultaneous display of illustrations (by means of a video tape device coupled with the cathode ray output tape) with the text matter related to them.

The development of this very high-speed system will permit the use of graphical materials in publications that are now being composed on line printers with no sacrifice in overall speed.

In 1967 this system will be employed by the US Air Force with these projected objectives:
reduction in 'response time' from a typical 28 days to 8 days,
lower manpower requirements of 40 man years,
reduced shipping costs: $451,000/year,
reduced rental costs of EDP equipment: $25,000/year,
reduced printing costs: $1,214,000/year.

Pages printed will go down from 1,053,000,000 to 446,000,000 per year. Legibility will be improved and bulk reduced. Interestingly, not one compositor will be affected since all of the work involved was hitherto composed on line printers.